IN THE CLAIMS

Please amend claim 1 as follows:

1.   (Currently Amended) A system for estimating the prevalence of digital content on a network, comprising:

an estimating device that receives traffic data collected from the network;

an anonymizing device that locates user identification data in the traffic data, masks the user identification data to produce clean traffic data, and stores the clean traffic data;

a sampling device that stores summarization data that ~~describe~~ describes each occurrence of the digital content in the clean traffic data; and

an accessing device that presents the clean traffic data and the summarization data to a user.

2.   (Previously Presented) The system of claim 1, wherein the estimating device receives the traffic data from at least one proxy cache server.

3.   (Previously Presented) The system of claim 1, wherein the sampling device computes the number of impressions of the digital content for a web site on the network.

4.   (Previously Presented) The system of claim 1, wherein the sampling device includes:

a prober that fetches a web page from the network;

an extractor that locates a fragment of the web page that includes the digital content; and

a classifier that performs a structural analysis of the fragment to classify the digital content.

5.   (Previously Presented) The system of claim 1, wherein the accessing device generates a report when the clean traffic data or the summarization data satisfy at least one criterion.

6.   (Previously Presented) A method of estimating the prevalence of digital content on a network, comprising the steps of:

estimating the global traffic to at least one Web site on the network to provide traffic data;

locating user identification data in the traffic data;

masking the user identification data to produce clean traffic data;

statistically sampling the contents of said at least one Web site to provide sampling data;

storing the clean traffic data and the sampling data; and

accessing the clean traffic data and the sampling data to generate a report.

7.   (Previously Presented) A system for estimating the prevalence of digital

content on a network connected to at least one network site that includes at least one network

server to access at least one uniform resource locator, the system comprising:

a database;

a traffic analysis system that receives traffic data from a traffic sampling

system, locates user identification data in the traffic data, masks the user identification data to

produce clean traffic data, and stores the clean traffic data in the database, the traffic data

including said at least one uniform resource locator;

a digital content sampling system that stores the digital content at said at

least one uniform resource locator in the database; and

a statistical summarization system that stores summarization data that

describe the digital content in the database.

8.   (Previously Presented) The system of claim 7, further comprising:

a Web front end that connects to the network and the database, wherein a

client uses a browser to connect to the Web front end.

9.   (Previously Presented) The system of claim 7, further comprising:

4

a user interface that an account manager, an operator, or a media editor

can use to administer the system.

10.   (Original) The system of claim 7, wherein the network is the Internet, and

wherein the network site is a Web site.

11.   (Previously Presented) The system of claim 7, wherein to mask the user

identification data in the traffic data the traffic analysis system replaces the user identification

data with a result from processing the user identification data through a cryptographically secure

one-way hash function.

12.   (Previously Presented) The system of claim 11, wherein the user

identification data includes a network address or a cookie.

13.   (Previously Presented) The system of claim 11, wherein the

summarization data includes a reference to said at least one uniform resource locator and a count

of the number of requests for said at least one uniform resource locator.

14.   (Previously Presented) The system of claim 7, wherein the digital content

sampling system further comprises:

a probemapping system that uses the summarization data to create a probe

map for the network, the probe map including a mapping for said at least one uniform resource

locator;

a uniform resource locator retrieval system that retrieves said at least one uniform resource locator from the network server;

a browser emulation environment that conducts a simulation of the display of said at least one uniform resource locator in a browser;

a digital content extractor that stores the digital content from said at least one uniform resource locator in the database; and

a structural classifier that stores at least one classification type for the digital content in the database.

15.   (Previously Presented) The system of claim 14, wherein the probe map further comprises:

a probability of the likelihood that said at least one uniform resource location will be sampled; and

a scale that determines the contribution of said at least one uniform resource location to the summarization data.

16.   (Original) The system of claim 14, wherein the simulation includes executing a program embedded in said at least one uniform resource locator.

17.   (Previously Presented) The system of claim 16, wherein the program is a
JavaScript script, a Java applet, a Perl script, or a common gateway interface program.

18.   (Original) The system of claim 14, wherein the simulation includes
executing dynamic digital content in said at least one uniform resource locator.

19.   (Previously Presented) The system of claim 18, wherein the dynamic
content is an interlaced GIF image, an MPEG movie, or an MP3 audio file.

20.   (Previously Presented) The system of claim 14, wherein the digital content
extractor retrieves the digital content from said at least one uniform resource locator by applying
a rule set defined by a media editor.

21.   (Previously Presented) The system of claim 14, wherein the digital content
extractor retrieves the digital content from said at least one uniform resource locator by using an
automated digital content detection system.

22.   (Previously Presented) The system of claim 21, wherein the automatic
digital detection system comprises:

        a structural detector that locates an XML structure; and

        a feature detector that locates an XML feature within the XML structure.

23.   (Previously Presented) The system of claim 14, wherein the structural

classifier determines said at least one classification type for the digital content.

24.   (Previously Presented) The system of claim 7, wherein the user interface

further comprises:

a system account management interface that assists the account manager

with creating and modifying an account on the system;

a site administration interface that assists the operator with the

administration of said at least one network site;

a taxonomy administration interface that assists the media editor with the

administration of the taxonomy data;

a digital content classification interface that assists the media editor with

the classification of the digital content; and

a rate card collection interface that assists the media editor with the

administration of the rate card data.

25.   (Previously Presented) A system for estimating prevalence of digital

content on a network, comprising:

a memory device; and

8

a processor disposed in communication with the memory device, the

processor configured to:

obtain traffic data from at least one Web site on the network;

locate user identification data in the traffic data;

mask the user identification data to produce clean traffic data;

compute a number of impressions for the digital content in the

clean traffic data;

retrieve the digital content from the clean traffic data to generate

sampling data; and

generate prevalence estimates for the digital content from the clean

traffic data and the sampling data.

     26.   (Previously Presented) The system of claim 25, wherein the processor is

further configured to:

retrieve a Web page from said at least one Web site;

extract a fragment from the Web page; and

classify the fragment.

27.   (Previously Presented) The system of claim 25, wherein to mask the user

identification data in the traffic data the processor is further configured to:

replace the user identification data with a result from processing the user

identification data through a cryptographically secure one-way hash function.

28.   (Previously Presented) The system of claim 27, wherein the user

identification includes a network address or a cookie.

29.   (Previously Presented) The system of claim 25, wherein the processor is

further configured to:

classify a fragment within the sampling data.

30.   (Previously Presented) The system of claim 29, wherein the processor is

further configured to:

classify the fragment by analyzing the fragment for uniqueness and adding

information to a database regarding the uniqueness of the fragment.

31.   (Previously Presented) The system of claim 30, wherein the processor is

configured to:

classify the fragment by detecting a duplicate fragment.

32.   (Previously Presented) The system of claim 25, wherein the processor is further configured to:

interact with a user interface that administers the system.

33.   (Previously Presented) The system of claim 25, wherein the processor is further configured to:

include uniform resource locator information regarding said at least one Web site in the traffic data.

34.   (Previously Presented) The system of claim 25, wherein the processor is further configured to:

perform data integrity monitoring of the sampling data.

35.   (Previously Presented) The system of claim 25, wherein the processor is further configured to:

serve as an automatic digital content detection system.

36.   (Previously Presented) The system of claim 35, wherein the automatic advertisement detection system applies at least one heuristic algorithm to detect digital content within an HTML or an XML document and normalizes the detected HTML or XML content into a hierarchical form.

37.   (Previously Presented) A method for using a computer to estimate the prevalence of digital content on a network, comprising the steps of:

obtaining traffic data from at least one Web site on the network;

locating user identification data in the traffic data;

masking the user identification data to produce clean traffic data;

computing a number of impressions for the digital content in the clean traffic data;

retrieving the digital content from the clean traffic data to generate sampling data; and

generating prevalence estimates for the digital content from the clean traffic data and the sampling data.

38.   (Previously Presented) The method of claim 37, wherein retrieving the digital content further comprises the steps of:

retrieving a Web page from said at least one Web site;

extracting a fragment from the Web page; and

classifying the fragment.

39.   (Previously Presented) The method of claim 37, wherein the masking of

the user identification data in the traffic data further comprises:

replacing the user identification data with a result from processing the user

identification data through a cryptographically secure one-way hash function.

40.   (Previously Presented) The method of claim 39, wherein the user

identification includes a network address or a cookie.

41.   (Previously Presented) The method of claim 37, further comprising the

step of:

classifying a fragment within the sampling data.

42.   (Previously Presented) The method of claim 41, wherein classifying the

fragment further comprises the steps of:

analyzing fragment for uniqueness; and

adding information to a database regarding the uniqueness of the

fragment.

43.   (Previously Presented) The method of claim 42, further comprising the

step of:

classifying the fragment by detecting a duplicate fragment.

44.   (Previously Presented) The method of claim 37, further comprising the

step of:

interacting with a user interface that administers the system.

45.   (Previously Presented) The method of claim 37, further comprising the

step of:

including uniform resource locator information regarding said at least one

Web site in the traffic data.

46.   (Previously Presented) The method of claim 37, further comprising the

step of:

performing data integrity monitoring of the sampling data.

47.   (Previously Presented) The method of claim 37, further comprising the

steps of:

performing automatic advertisement detection by applying at least one

heuristic algorithm to detect advertising within an HTML or an XML document; and

normalizing the detected HTML or XML content into a hierarchical form.

48.   (Previously Presented) A computer readable medium comprising:

code for obtaining traffic data from at least one Web site on the network;

code for locating user identification data in the traffic data;

code for masking the user identification data to produce clean traffic data;

code for computing a number of impressions of digital content in the clean

traffic data;

code for retrieving the digital content from the clean traffic data to

generate sampling data; and

code for generating prevalence estimates for the digital content from the

clean traffic data and the sampling data.

49. (Previously Presented) The computer readable medium of claim 48,

further comprising:

code for retrieving a Web page from said at least one Web site; code for

extracting a fragment from the Web page; and code to classify the fragment.

50. (Previously Presented) A system for estimating prevalence of digital

content on a network, comprising:

means for obtaining traffic data from at least one Web site on the network;

means for locating user identification data in the traffic data;

means for masking the user identification data to produce clean traffic

data:

means for computing a number of impressions for the digital content in the

clean traffic data;

means for retrieving the digital content from the clean traffic data to

generate sampling data; and

means for generating prevalence estimates of the digital content from the

clean traffic data and the sampling data.

51.   (Previously Presented) The system of claim 50, further comprising:

means for classifying a fragment extracted from a Web page.

52.   (Previously Presented) The system of claim 50, further comprising:

means for replacing the user identification data with a result from

processing the user identification data through a cryptographically secure one-way hash function.

53.   (Previously Presented) A system of estimating prevalence of digital

content on a network, comprising:

means for estimating global traffic to at least one Web site on the network

to provide traffic data;

means for locating user identification data in the traffic data;

means for masking the user identification data to produce clean traffic
data:

means for statistically sampling the contents of said at least one Web site
to provide sampling data;

means for storing the clean traffic data and the sampling data; and

means for generating prevalence estimates for the digital content by
accessing the clean traffic data and the sampling data.

54.   (Previously Presented) The system of claim 53, further comprising:

means for reporting the prevalence estimates to a user.

55.   (Previously Presented) A method for using a computer to estimate the
prevalence of digital content on a network, comprising the steps of:

receiving traffic data from the network:

locating user identification data in the traffic data;

masking the user identification data to produce clean traffic data;

storing the clean traffic data;

17

storing summarization data that describe each occurrence of the digital

content in the clean traffic data; and

presenting the clean traffic data and the summarization data to a user.

56.   (Previously Presented) The method of claim 55, wherein the receiving of

the traffic data is from at least one proxy server.

57.   (Previously Presented) The method of claim 55, wherein storing

summarized traffic data further comprises the step of:

computing the number of impressions of the digital content for a web site

on the network.

58.   (Previously Presented) The method of claim 55, wherein storing traffic

data further comprises the steps of:

fetching a web page from the network;

locating a fragment of the web page that includes the digital content; and

performing a structural analysis of the fragment to classify the digital

content.

59.   (Previously Presented) The method of claim 55, wherein presenting the

clean traffic data and the summarization data further comprises the step of:

18

generating a report when the clean traffic data or the summarization data

satisfy at least one criterion.

60.   (Previously Presented) A system for estimating prevalence of digital

content on a network, comprising:

a memory device; and

a processor disposed in communication with the memory device, the

processor configured to:

receive traffic data from the network:

locate user identification data in the traffic data;

mask the user identification data to produce clean traffic data;

store the clean traffic data;

store summarization data that describe each occurrence of the

digital content in the clean traffic data; and

present the clean traffic data and the summarization data to a user.

61.   (Previously Presented) The system of claim 60, wherein the receiving of

the traffic data is from at least one proxy server.

62.   (Previously Presented) The system of claim 60, wherein the processor

computes the number of impressions of the digital content for a web site on the network.

63.   (Previously Presented) The system of claim 60, wherein the processor is

further configured to:

fetch a web page from the network;

locate a fragment of the web page that includes the digital content; and

perform a structural analysis of the fragment to classify the digital content.

64.   (Previously Presented) The system of claim 60, wherein the processor

generates a report when the clean traffic data or the summarization data satisfy at least one

criterion.

65.   (Previously Presented) A computer readable medium comprising:

code for receiving traffic data from the network;

code for locating user identification data in the traffic data;

code for masking the user identification data to produce clean traffic data;

code for storing the clean traffic data;

code for storing summarization data that describe each occurrence of the

digital content in the clean traffic data; and

code for presenting the clean traffic data and the summarization data to a

user.

66.   (Previously Presented) The computer readable medium of claim 65, the

receiving of the traffic data is from at least one proxy server.

67.   (Previously Presented) The computer readable medium of claim 65,

further comprising:

code for computing the number of impressions of the digital content for a

web site on the network.

68.   (Previously Presented) The computer readable medium of claim 65,

further comprising:

code for fetching a web page from the network;

code for locating a fragment of the web page that includes the digital

content; and code for performing a structural analysis of the fragment to classify the digital

content.

69.   (Previously Presented) The computer readable medium of claim 65,

further comprising:

code for generating a report when the clean traffic data or the

summarization data satisfy at least one criterion.